Public Quarterly Report

Date of Report: 3rd Quarterly Report-June 29, 2020

Contract Number: 693JK31910002POTA
Prepared for: DOT PHMSA
Project Title: Data Collection, Normalization and Integration Methods to Enhance Risk Assessment Tools for Decision-Making
Prepared by: GTI (Gas Technology Institute)
Contact Information: Ernest Lever, elever@gti.energy , 847-544-3415
For quarterly period ending: June 29th, 2020

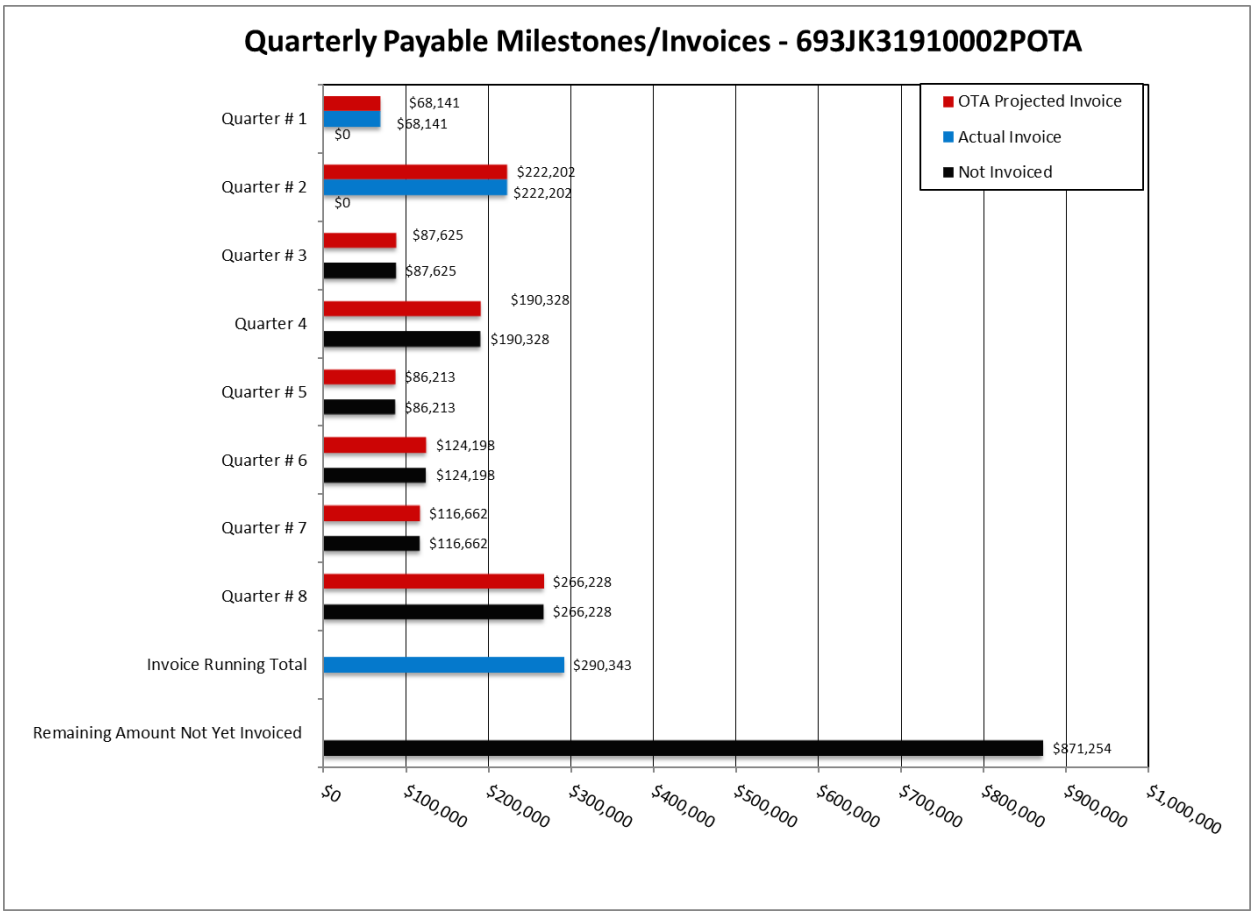
1: Items Completed During this Quarterly Period:

Note: Deliverables are those scheduled for the 2nd quarterly report in the contract. A contract modification will be submitted in July 2020 to properly align the deliverables with the extended period of execution that was approved by PHMSA

Item #	*Task #*	*Activity/Deliverable*	*Title*	*Federal Cost*	*Cost Share*
6	2.1	Data Sources and Methods	State-of-the-art Review	19,194.00	0.00
7	3.1	Decision Support System	Prototype decision support system design, screen mockups	26,974.00	27,500.00
8	4.1.1	Data Integration, Normalization and Uncertainty Tools	Bayesian Network: Node definitions	48,964.00	0.00
9	4.1.2	Data Integration, Normalization and Uncertainty Tools	Bayesian Network: Arc definitions	48,964.00	0.00
10	4.2.2	Data Integration, Normalization and Uncertainty Tools	Data Driven Insights: Identify High Impact Decision Relevant Questions	10,980.00	21,746.00
11	4.3.1	Data Integration, Normalization and Uncertainty Tools	Causal Modeling and Situational Awareness: Causal modeling framework Progress report 2	19,194.00	0.00

Item #	*Task #*	*Activity/Deliverable*	*Title*	*Federal Cost*	*Cost Share*
12	4.3.2	Data Integration, Normalization and Uncertainty Tools	Causal Modeling and Situational Awareness: Situational awareness framework Progress report 1	42,813.00	0.00
13	5.1	Project Management	Second Quarterly Report	5,119.00	0.00

3: Project Financial Tracking During this Quarterly Period:



4: Project Technical Status –

The technical work for this quarter progressed as planned and all deliverables have been completed. Detailed documentation for each deliverable is provided in attachments. A short description of project activity is provided in the following numbered points:

1. The state-of-the-art review identified industry best practices for addressing data quality. The review included
    a. The body of knowledge organized by DAMA International, the global data management community,

b. International standards pertaining to data quality and data governance,
c. Frameworks for record keeping and control to identify quality concerns related to record keeping
d. Books and articles published be practicing data management professionals
e. Standards for business model and notation, and decision model and notation
f. Enterprise governance frameworks
g. Data governance approaches
h. Risk governance frameworks
i. Literature on information fusion that address
    i. Information quality
    ii. Information fusion
    iii. Decision making
    iv. Context enhanced information fusion
j. Situational awareness frameworks

2. The Lumina task related to decision support systems covered:
    a. A general framework for risk decision making, presented as an influence diagram that provides the central organizing architecture for the decision-support tool.
    b. A prototype causal influence diagram to estimate leak probabilities based on characteristics of pipeline assets.
    c. A prototype causal influence diagram and model to estimate the risks of rupture of distribution lines from overpressure due to incorrect operations (IO).
    d. Design of a tool to import and incorporate risk models into ANAGRAM, including Bayesian networks using the Genie software, being developed by UMD and GTI.
    e. Methods to calculate Risk Spend Efficiency (RSE) as a metric to compare mitigation projects in terms of the ratio of reduction in risk to the cost of mitigation.
    f. Extensions to evaluate not just individual mitigation projects, but also portfolios of projects.
    g. Methods to scale computations to extend from transmission to distribution networks, which typically have 5 to 10 times as many assets.
    h. Design and mock-ups of user interface screens to show how an enhanced and extended version of ANAGRAM can enable gas utility risk managers and project planners to make use of the elements listed above.

3. THE UMD team performed a comprehensive review of previous work on locates and third-party damage performed by GTI. Their review extended to available data sources and their technical deliverables include:
    a. An underlying node structure was developed for a comprehensive third-party damage BN.
    b. The nodes and arc definitions were developed using varying techniques that emphasized the importance of key features to come out through the research.
    c. Preliminary work on analyzing and understanding sponsored TPD data by PG&E also occurred, offering insight into the current damage reporting practices and the opportunities and challenges that they show in terms of quantifying and expanding the BN model presented in this report.
    d. Next steps include two tasks:
        i. Process the data and parameter learning. In the data processing step, we will be mapping the provided datasets from the utility partner PG&E onto the nodes included in the BN model. This data will be used to further improve the BN model causal structure, incorporating interdependencies found between different TPD causes.
        ii. In addition, public databases such as PHMSA incident reports and CGA DIRT reports, among others, are going to be identified and analyzed to further improve this. Following the data mapping, we will conduct parameter learning to populate the BN node probability tables from appropriate quantitative data sets
4. The Stanford work in this quarter is summarized as follows:
    a. In the first three months, we held extensive discussions with PG&E and other stakeholders to identify key datasets and important general areas of focus that these data sets would allow us to address.
    b. Over the past three months, we have received and begun to analyze many of the most important data sets, made progress toward acquiring supplementary data sets, and solidified well-scoped, decision-relevant research questions that we can pursue in coming quarters.
    c. Our preliminary analysis of several years of data for essentially the entire population of PG&E distribution assets suggests that the data set is indeed well-suited to addressing the predictive (machine learning) and causal (econometric) questions of predicting and quantifying leakage risk.

d. The fact that we have leak data for the entire PG&E service territory will help us overcome inevitable difficulties related to predicting and otherwise statistically analyzing what are thankfully relatively low-probability events within a given city.

e. The addition of expansive third-party data sources such as satellite based InSAR will allow us to determine the value of these tools for leakage prediction and may allow us to estimate the causal effects of ground movement on distribution system pipeline integrity using econometric analysis.

f. With this approach, we can also assess the value of higher-quality information provided commercially available InSAR satellite networks over that of publicly available networks.

g. Next steps will address the following:

    i. We will continue our efforts to obtain vehicle-based air quality surveys

        1. If we can use data from recurring vehicle-based methane, ethane, and air quality surveys, we will further be able to evaluate the predictive capabilities of such tools in proactively identifying leaks.

        2. These data will also likely provide insight into leak persistence.

    ii. Next quarter, we will begin detailed analysis of our data and test the predictive capabilities of the data using various machine learning models.

        1. We will integrate the results of these predictive and causal analyses into the larger holistic decision support framework being developed in collaboration with GTI, Lumina Decision Systems, and University of Maryland.

        2. This will place the results of the leakage and subsidence analysis in the larger context of the suite of risks faced by utility integrity management planners, allowing them to more effectively mitigate risk with limited resources.

5. Causal modeling Framework Progress Report 2

    a. o date we have provided two examples of causal models for gas distribution systems that deal with a complex set of interacting factors that drive system failure:

b. In the previous quarterly report, a model for vintage Aldyl A pipelines was presented. The model captures the physical process driving the slow-crack-growth mechanism as well as external factors that influence the failure mechanism. The model was run on several actual distribution systems that have widely differing performance characteristics and it was shown how the model can be used to both calculate exact expected performance of a system and how to compare the expected performance to well vetted reference data. It was also shown how to use the causal modeling framework to run multiple scenarios that explore the influence of the various parameters on system performance.

c. In this report UMD explain a preliminary causal model that provides a comprehensive framework for addressing the interacting factors that contribute to the likelihood of third-party damage to a distribution system

d. The state-of-the-art review reported on this quarter provides a comprehensive justification as to why causal modeling frameworks are well suited to address the complexities of modeling risk related to natural gas infrastructure.

6. Situational Awareness Framework Progress report 1

   a. A brief overview is provided as to how the situational awareness frameworks ties into the overall approach

   b. Detailed information is provided on how to develop a cognitive map through Goal Directed Task Analysis

5: Project Schedule

a. The project is on schedule and on budget

<u>End of report summary</u>